Submitted to Rev. Mat. Iberoam., 1–24

The dual conjecture of Muckenhoupt and Wheeden

Adam Osękowski

Abstract.
Let T be a Calderón-Zygmund operator on $\mathbb{R}^d$. We prove the existence of a constant $C_{T,d} < \infty$ such that for any weight w on $\mathbb{R}^d$ satisfying Muckenhoupt's condition A_1, we have

$$w\big(\{x \in \mathbb{R}^d : |Tf(x)| > w(x)\}\big) \leq C_{T,d}[w]_{A_1} \int_{\mathbb{R}^d} f \mathrm{d}x.$$

The linear dependence on $[w]_{A_1}$, the A_1 characteristic of w, is optimal. The proof exploits the associated dimension-free inequalities for dyadic shifts, which are of independent interest.

1. Introduction

Let M be the Hardy-Littlewood maximal operator, acting on locally integrable functions on $\mathbb{R}^d$ by the formula

$$Mf(x) = \sup \Big\{ \langle f \rangle_Q \chi_Q(x) \Big\},$$

where the supremum is taken over all cubes $Q \subset \mathbb{R}^d$ with sides parallel to the axes, and $\langle f \rangle_Q = \frac{1}{|Q|} \int_Q f \mathrm{d}x$ denotes the average of f over Q. A celebrated result of Fefferman and Stein [7] established in 1971 asserts that if w is an arbitrary weight on $\mathbb{R}^d$, i.e., a nonnegative, locally integrable function, then

$$w\big(\{x \in \mathbb{R}^d : Mf(x) \geq 1\}\big) \leq C_d \int_{\mathbb{R}^d} |f| Mw \mathrm{d}x.$$

Here we use the notation $w(E) = \int_E w \mathrm{d}x$ for the measure associated with w and the constant C_d depends only on the dimension d. A few years later Muckenhoupt and Wheeden conjectured that a similar weak-type bound holds true for an arbitrary

Mathematics Subject Classification (2010): Primary 42B25; Secondary 46E30.
Keywords: Dyadic, shift, weight, Bellman function, best constant.

Calderón-Zygmund singular integral operator T: there is a finite constant $C_{T,d}$ depending only on the parameters indicated such that for any f and any weight w,

$$w\big(\{x\in\mathbb{R}^d : Tf(x)\geq 1\}\big)\leq C_{T,d}\int_{\mathbb{R}^d}|f|Mw\mathrm{d}x. \tag{1.1}$$

There is a weaker version of the conjecture for the so-called A_1 weights. Recall that w satisfies Muckenhoupt's condition A_1 (or belongs to the class A_1) if there is a finite constant K such that $Mw\leq Kw$ almost everywhere; the smallest K with this property is denoted by $[w]_{A_1}$ and called the A_1 characteristics of w. If (1.1) held true, it would imply that any A_1 weight w satisfies

$$w\big(\{x\in\mathbb{R}^d : Tf(x)\geq 1\}\big)\leq c[w]_{A_1}\int_{\mathbb{R}^d}|f|w\mathrm{d}x, \tag{1.2}$$

which is referred to as the weak conjecture of Muckenhoupt and Wheeden. Both (1.1) and (1.2) turned out to be very difficult questions and were studied from different perspectives (e.g., for special operators T or special weights w): consult e.g. [1, 5, 13, 14, 22]. In 2010, both conjectures were finally shown to be false: they do not hold true even for $d=1$ and T being the Hilbert transform. See the counterexamples by Reguera [23] and Reguera and Thiele [24], consult also the works of Nazarov, Reznikov, Vasyunin and Volberg [15, 16]. In the weak conjecture, it turns out that an additional logarithmic factor is required. As shown in [14], we have the estimate

$$w\big(\{x\in\mathbb{R}^d : Tf(x)\geq 1\}\big)\leq c[w]_{A_1}\log([w]_{A_1}+e)\int_{\mathbb{R}^d}|f|w\mathrm{d}x,$$

and the LlogL-dependence on $[w]_{A_1}$ is optimal (cf. [11]).

There are dual versions of the estimates (1.1) and (1.2), which appeared for the first time in the work of Lerner, Ombrosi and Pérez [12]. The strong version is

$$w\left(\{x\in\mathbb{R}^d : |Tf(x)|\geq Mw(x)\}\right)\leq C_{T,d}\int_{\mathbb{R}^d}|f|\mathrm{d}x, \tag{1.3}$$

where w is an arbitrary weight, T is a Calderón-Zygmund operator and $C_{T,d}$ depends only on T and the dimension. The weaker inequality concerns A_1 weights and reads

$$w\left(\{x\in\mathbb{R}^d : |Tf(x)|\geq w(x)\}\right)\leq C_{T,d}[w]_{A_1}\int_{\mathbb{R}^d}|f|\mathrm{d}x. \tag{1.4}$$

To understand why these can be regarded as dual statements, suppose that (1.1) holds true for some T and apply the extrapolation theorem of Cruz-Uribe and Pérez [6]. Then for any $1<p<\infty$ there is a constant $c=c_{T,d,p}$ such that

$$\int_{\mathbb{R}^d}|Tf|^p w\mathrm{d}x\leq c_{T,d,p}\int_{\mathbb{R}^d}|f|^p\left(\frac{Mw}{w}\right)^p w\mathrm{d}x$$

for all f and w. By duality, this estimate yields

$$\int_{\mathbb{R}^d} \left(\frac{|T^*f|}{Mw}\right)^{p'} w\mathrm{d}x \le C \int_{\mathbb{R}^d} \left(\frac{|f|}{w}\right)^{p'} w\mathrm{d}x,$$

where $p' = p/(p-1) \in (1,\infty)$. Thus (1.3) can be regarded as a limiting weak-type (1,1) version, as $p \to \infty$, of this estimate (applied to the operator T^*), and the restriction to A_1 weights yields (1.4).

In [12], Lerner, Ombrosi and Pérez proved the following weaker form of (1.3): there is a constant c depending only on n and T such that

$$w\left(\{x \in \mathbb{R}^n : |Tf(x)| \ge M^3 w(x)\}\right) \le c \int_{\mathbb{R}^n} |f|\mathrm{d}x$$

(here M^3 is the third iteration of the Hardy-Littlewood maximal operator). In [21], it was proved that (1.3) does not hold in the special dyadic case of Haar multipliers. It can be extracted from the construction in [24] that (1.3) fails as well, see also [4] for an extension of this fact (and consult the Appendix at the end of this paper). Concerning (1.4), we will prove the following statement.

Theorem 1.1. *The weak conjecture* (1.4) *holds true for general Calderón-Zygmund operators and the linear dependence on the characteristic* $[w]_{A_1}$ *is already the best possible for the Hilbert transform.*

This is quite surprising in the light of (1.1) and (1.2): the strong dual conjecture fails, while the weak counterpart is true. Our approach will rest on dyadic approximations and, in particular, will make heavy use of dyadic shifts and sparse operators. Let us present the necessary background on this subject: see the recent work [10] of Lerner and Nazarov for a detailed exposition. Suppose that $\mathscr{D}$ is a dyadic lattice in $\mathbb{R}^d$ (not necessarily the standard one), i.e., a collection of cubes satisfying the following three conditions:

(i) if $Q \in \mathscr{D}$, then each child of Q is in $\mathscr{D}$,
(ii) every two cubes Q' , $Q'' \in \mathscr{D}$ have a common ancestor.
(iii) for every compact set $K \in \mathbb{R}^d$, there exists a cube $Q \in \mathscr{D}$ containing K.

A sequence $\alpha = (\alpha_Q)_{Q\in\mathscr{D}}$ of nonnegative numbers has Carleson property, if there exists a constant $\eta > 0$ such that

$$\sup_{R\in\mathscr{D}} \frac{1}{|R|} \sum_{Q\in\mathscr{D}, Q\subseteq R} \alpha_Q |Q| \le \eta,$$

where $|\cdot|$ is the Lebesgue measure. Sometimes, to indicate the value of the constant η, a sequence α with the above property will be called η-Carleson. It is well-known that if α is η-Carleson, then there exists a family $\{E(Q)\}_{Q\in\mathscr{D}}$ of pairwise disjoint sets such that $E(Q) \subseteq Q$ and $\alpha_Q = \eta|E(Q)|/|Q|$ for each Q. For any Carleson sequence α, we introduce the associated shift operator $\mathcal{A}_\alpha$ acting on locally integrable functions $f : \mathbb{R}^d \to \mathbb{R}$ by

$$\mathcal{A}_\alpha f = \sum_{Q\in\mathscr{D}} \alpha_Q \langle f\rangle_Q \chi_Q. \tag{1.5}$$

Such objects are closely related to the class of the so-called sparse operators. Recall that a collection $\mathcal{S} \subset \mathscr{D}$ is called sparse, if there is a family $\{E(Q)\}_{Q\in\mathcal{S}}$ of pairwise disjoint sets such that $E(Q) \subseteq Q$ and $|E(Q)| \geq |Q|/2$ for each $Q \in \mathcal{S}$. Given any such class $\mathcal{S}$, one defines the associated sparse operator by

$$\mathcal{E}_{\mathcal{S}} f = \sum_{Q\in\mathcal{S}} \langle f\rangle_Q \chi_Q.$$

Such an operator is of the form (1.5), if one sets $\alpha_Q = 1$ for $Q \in \mathcal{S}$ and $\alpha_Q = 0$ otherwise; in addition, it is easy to see that the sparseness of $\mathcal{S}$ implies that the sequence $(\alpha_Q)_{Q\in\mathscr{D}}$ is $2-$Carleson.

One of the important features of the sparse operators (and hence also the class of shift operators) lies in the fact that these objects dominate, in an appropriate sense, large families of Calderón-Zygmund singular integrals [9, 10]: this will also be formulated precisely in Section 4 below. To establish Theorem 1.1, we will prove first its dyadic counterpart, which can be stated as follows.

Theorem 1.2. *Suppose that $\mathscr{D}$ is an arbitrary dyadic lattice in $\mathbb{R}^d$ and $\alpha = (\alpha_Q)_{Q\in\mathscr{D}}$ is an η-Carleson sequence. Then for any integrable function f on $\mathbb{R}^d$ and any A_1 weight w on $\mathbb{R}^d$ we have*

$$w(|\mathcal{A}_\alpha f| \geq w) \leq 8\eta[w]_{A_1}\|f\|_{L^1(\mathbb{R}^d)}. \tag{1.6}$$

The linear dependence on the A_1 characteristic is the best possible.

The nice feature of this result is that the multiplicative constant 8 does not depend on the dimension. Actually, we will prove the above statement in a much more general setting. The regularity and self-similarity of the dyadic lattice is not needed, we will study the estimate in the context of probability measures equipped with a tree-like structure. Here is the precise definition.

Definition 1.3. *Suppose that (X, μ) is a nonatomic probability space. A set $\mathcal{T}$ of measurable subsets of X will be called a tree if the following conditions are satisfied:*

- *(i) $X \in \mathcal{T}$ and for every $Q \in \mathcal{T}$ we have $\mu(Q) > 0$.*
- *(ii) For every $Q \in \mathcal{T}$ there is a finite subset $C(Q) \subset \mathcal{T}$ containing at least two elements such that*
  - *(a) the elements of $C(Q)$ are pairwise disjoint subsets of Q,*
  - *(b) $Q = \bigcup C(Q)$.*
- *(iii) $\mathcal{T} = \bigcup_{m\geq 0} \mathcal{T}^m$, where $\mathcal{T}^0 = \{X\}$ and $T^{m+1} = \bigcup_{Q\in\mathcal{T}^m} C(Q)$.*
- *(iv) We have $\lim_{m\to\infty} \sup_{Q\in\mathcal{T}^m} \mu(Q) = 0$.*

All the objects introduced above in the dyadic setting can be generalized to the probabilistic context, simply by replacing $\mathscr{D}$ with $\mathcal{T}$ and $(\mathbb{R}^d, |\cdot|)$ with (X, μ). The only slight change concerns the A_1 condition, in which one must use the maximal operator $M^{\mathcal{T}}$ associated with $\mathcal{T}$:

$$M^{\mathcal{T}} f = \sup_{Q\in\mathcal{T}} \Big(\langle f\rangle_{Q,\mu}\chi_Q\Big).$$

We will prove the following.

Theorem 1.4. *Let (X,μ) be a probability space with a tree structure $\mathcal{T}$. Assume further that $\alpha = (\alpha_Q)_{Q\in\mathcal{T}}$ is an η-Carleson sequence. Then for any integrable function f on X and any A_1 weight w on X we have*

$$w(|\mathcal{A}_\alpha f| \geq w) \leq 8\eta [w]_{A_1} \|f\|_{L^1(X)}. \tag{1.7}$$

The linear dependence on the A_1 characteristic is the best possible.

Let us stress here that we do not impose any regularity condition on $\mathcal{T}$: for any element Q of $\mathcal{T}$ and any child Q' of Q, the ratio $\mu(Q')/\mu(Q)$ need not be bounded away from 0 or 1. It is easy to see that the above result is an extension of Theorem 1.2. Indeed, given a dyadic lattice $\mathscr{D}$, we pick an arbitrary base cube $Q \in \mathscr{D}$ and consider the probability space $(Q, |\cdot|/|Q|)$ equipped with the dyadic tree. Now, any A_1 weight w on $\mathbb{R}^d$, when restricted to Q, becomes the probabilistic weight with the characteristic less or equal to $[w]_{A_1}$ and hence (1.7) holds true. Multiplying both sides by $|Q|$ and letting $|Q| \to \infty$ gives (1.6). We would also like to mention here that the estimate (1.7), and hence (1.6) as well, holds true for A_p weights (for $p > 1$), see [3] and consult the references therein. However, it is not clear whether the constant involved depends linearly on $[w]_{A_p}$.

A few words about the proof of the inequality (1.7) are in order. Our approach will make use of Bellman function method, a powerful tool used widely in harmonic analysis and probability theory. This technique has its origins in the theory of optimal stochastic control, and its connections with other areas of mathematics were firstly observed by Burkholder, who used it to identify the unconditional constants of the Haar system. Soon after the appearance of [2], Burkholder's arguments were extended by a number of mathematicians to investigate numerous estimates for semimartingales: see e.g. the monograph [19] for an overview. In the nineties, the seminal paper [17] by Nazarov and Treil (inspired by the preprint version of [18]) pushed the technique towards applications in harmonic analysis; since then, the method has been used in many contexts, including BMO inequalities, weighted estimates and many more. Roughly speaking, the Bellman function method relates the validity of a given estimate to the existence of a certain special function which enjoys appropriate size and concavity conditions. Typically, this function is quite complicated and its discovery, as well as the verification of the required properties, is a quite elaborate issue. Our approach will enable us to overcome this difficulty: we will obtain an *abstract*, non-explicit formula for the Bellman function corresponding to (1.7). This argument was motivated by a similar phenomenon which occurs in the classical, well-understood context of Haar multipliers on the interval $[0,1]$. We strongly believe that this novel argument is applicable in a number of related results in the area.

The next section is devoted to the explanation of the above idea of obtaining abstract Bellman functions for weak-type estimates from the corresponding objects coming from L^2 estimates. Section 3 contains the detailed exposition of the Bellman function method in the context of shift operators $\mathcal{A}_\alpha$. In Section 4 we provide the proof of Theorem 1.4, and in Section 5 we establish the results for

Calderón-Zygmund operators formulated in Theorem 1.1. The final part of the paper contains the construction of the counterexample related to (1.3).

2. A motivating example

In this section we study a very simple context of Haar multipliers, following the works of Burkholder [2]. Let $(h_n)_{n\geq 0}$ be the standard Haar system on $[0,1)$, i.e., a collection of functions given by $h_0 = \chi_{[0,1)}$, $h_1 = \chi_{[0,1/2)} - \chi_{[1/2,1)}$, $h_2 = \chi_{[0,1/4)} - \chi_{[1/4,1/2)}$, $h_3 = \chi_{[0,1/2)} - \chi_{[1/2,1)}$, and so on. Suppose that $V : \mathbb{R}^2 \to \mathbb{R}$ is a given function and assume that we are interested in showing the inequality

$$\int_{[0,1)} V\left(\sum_{k=0}^n a_k h_k, \sum_{k=0}^n \epsilon_k a_k h_k\right) \mathrm{d}x \leq 0 \qquad n = 0,\, 1,\, 2,\, \ldots, \tag{2.1}$$

for any sequence $(a_k)_{k\geq 0}$ of integers and any sequence $(\epsilon_k)_{k\geq 0}$ of signs. For instance, for the choice $V(x,y) = |y|^p - C_p^p|x|^p$ (where $1 < p < \infty$) the above estimate is related to the unconditionality of the Haar system. The key to handle this problem is to consider the class of all functions $\mathcal{B} : \mathbb{R}^2 \to \mathbb{R}$ which enjoy the following properties:

1° (Initial condition) $\mathcal{B}(x, \pm x) \leq 0$ for all $x \in \mathbb{R}$;
2° (Majorization) $\mathcal{B} \geq V$ on $\mathbb{R}^2$;
3° (Concavity-type property) $\mathcal{B}$ is concave along any line of slope ± 1.

The existence of a function $\mathcal{B}$ with the above properties implies the validity of (2.1). Indeed, the third condition implies that for any $n \geq 0$ we have

$$\int_0^1 \mathcal{B}\left(\sum_{k=0}^{n+1} a_k h_k, \sum_{k=0}^{n+1} \epsilon_k a_k h_k\right) \mathrm{d}x \leq \int_0^1 \mathcal{B}\left(\sum_{k=0}^{n} a_k h_k, \sum_{k=0}^{n} \epsilon_k a_k h_k\right) \mathrm{d}x$$

(this is just the conditional Jensen's inequality), so by 2° and finally 1°, we obtain

$$\begin{aligned}\int_0^1 V\left(\sum_{k=0}^{n} a_k h_k, \sum_{k=0}^{n} \epsilon_k a_k h_k\right) \mathrm{d}x &\leq \int_0^1 \mathcal{B}\left(\sum_{k=0}^{n} a_k h_k, \sum_{k=0}^{n} \epsilon_k a_k h_k\right) \mathrm{d}x \\ &\leq \int_0^1 \mathcal{B}(a_0, \epsilon a_0) \mathrm{d}x \leq 0.\end{aligned}$$

Probably the simplest inequality which can be studied with the above approach is the L^2 bound

$$\left\|\sum_{k=0}^n \epsilon_k a_k h_k\right\|_{L^2}^2 \leq \left\|\sum_{k=0}^n a_k h_k\right\|_{L^2}^2,$$

$n = 0,\, 1,\, 2,\, \ldots$ (which, of course, follows at once from the orthogonality of the Haar system). The corresponding function V, i.e., the one which transforms the L^2 bound into (2.1), is given by $V(x,y) = y^2 - x^2$, and it turns out that $\mathcal{B} = V$

is the corresponding special function. Let us see what happens for the weak-type $(1,1)$ estimate

$$\left|\left\{x\in[0,1):\left|\sum_{k=0}^n \epsilon_k a_k h_k(x)\right|>1\right\}\right|\leq C\left\|\sum_{k=0}^n a_k h_k\right\|_{L^1},$$

for $n=0, 1, 2, \ldots$. This inequality is of the form (2.1), with $V(x,y)=\chi_{\{|y|>1\}}-C|x|$, and using the above approach, Burkholder showed the estimate with the optimal constant $C=2$. The special function $\mathcal{B}$ is slightly more complicated:

$$\mathcal{B}(x,y)=\begin{cases} y^2-x^2 & \text{if } |x|+|y|\leq 1,\\ 1-2|x| & \text{if } |x|+|y|>1.\end{cases}$$

For some more or less formal arguments which lead to the discovery of this function, see e.g. [19, 20]. For the sake of our further considerations concerning the estimate (1.7), let us make here some important observations. We see that $\mathcal{B}$ is built of two components: if (x,y) is close to $(0,0)$, then it coincides with the special function corresponding to the L^2 estimate; for remaining (x,y), it is an affine expression (in $|x|$), which is almost equal to V. One easily checks 1° and 2°; to verify 3°, we rewrite the above formula as

$$\mathcal{B}(x,y)=\begin{cases} \min\left\{y^2-x^2, 1-2|x|\right\} & \text{if } |x|\leq 1,\\ 1-2|x| & \text{if } |x|>1\end{cases} \tag{2.2}$$

and now it is clear that the concavity holds: both $(x,y)\mapsto y^2-x^2$ and $(x,y)\mapsto 1-2|x|$ are concave along the lines of slope ± 1, and hence so is $\mathcal{B}$, being essentially the minimum of the two.

As we will see in Section 4, the inequality (1.7) can be efficiently studied in a similar manner: it will be handled with a certain Bellman function given as the minimum of special functions associated with L^2 estimates and the appropriate affine expressions. More precisely, we will proceed as follows: first we will prove *directly* a certain weighted L^2 estimate for dyadic shifts; this will give us the existence of the associated Bellman function $\mathfrak{B}$. Then we will take an appropriate modification of the formula (2.2), with the term y^2-x^2 replaced with $\mathfrak{B}$, to obtain the function for the weak-type estimate.

3. Bellman function method for shift operators

We return to the context of arbitrary probability space (X,μ) equipped with a tree-like structure $\mathcal{T}$. Let $c\in[1,\infty)$ be a given parameter and let $V:[0,\infty)^3\to\mathbb{R}$ be a fixed function. Suppose further that we are interested in showing the estimate

$$\int_X V\left(f, \mathcal{A}_\alpha^{X,\mathcal{T}} f, w\right)\mathrm{d}\mu\leq 0 \tag{3.1}$$

for any integrable function $f : X \to [0,\infty)$, any 1-Carleson sequence $\alpha = (\alpha_Q)_{Q\in\mathcal{T}}$ (which gives rise to the corresponding shift $\mathcal{A}_\alpha^{X,\mathcal{T}}$) and any A_1 weight w on X satisfying $[w]_{A_1} \leq c$. Here the probability space (X,μ) and a tree structure $\mathcal{T}$ are also allowed to vary. To handle this problem, consider the five-dimensional domain

$$D = D_{p,r,c} = \left\{(x,y,u,v,t) \in [0,\infty)^4 \times [0,1] \,:\, c^{-1}v \leq u \leq v\right\}$$

and consider the class of special functions $\mathcal{B} : D \to \mathbb{R}$ which enjoy the following structural properties.

1° (Initial condition) We have

$$\mathcal{B}(x,0,u,v,t) \leq 0 \qquad \text{if } (x,0,u,v,t) \in D. \tag{3.2}$$

2° (Majorization) If $(x,y,u,v,t) \in D$, then we have the inequality

$$\mathcal{B}(x,y,u,v,0) \geq V(x,y,u). \tag{3.3}$$

3° (Monotonicity-type property) For any $(x,y,u,v,t) \in D$ and any $\delta \in [0,t]$,

$$\mathcal{B}(x,y,u,v,t) \geq \mathcal{B}(x,y+\delta x,u,v,t-\delta). \tag{3.4}$$

4° (Concavity-type property) For any $(x_\pm,y,u_\pm,v_\pm,t_\pm)$, $(x,y,u,v,t) \in D$ and any $\lambda \in [0,1]$ such that $x = \lambda x_- + (1-\lambda)x_+$, $u = \lambda u_- + (1-\lambda)u_+$, $t = \lambda t_- + (1-\lambda)t_+$ and $v_\pm = \max\{u_\pm, v\}$, we have

$$\mathcal{B}(x,y,u,v,t) \geq \lambda\mathcal{B}(x_-,y,u_-,v_-,t_-) + (1-\lambda)\mathcal{B}(x_+,y,u_+,v_+,t_+). \tag{3.5}$$

The relation between the existence of such a function $\mathcal{B}$ and the validity of (3.1) is studied in the two theorems below. We will need the following auxiliary fact, which combines the above conditions 3° and 4° into an enhanced concavity.

Lemma 3.1. *Suppose that $\mathcal{B} : D \to \mathbb{R}$ satisfies 3° and 4°. Then for any any integer $m \geq 2$, we have the estimate*

$$\mathcal{B}(x,y,u,v,t) \geq \sum_{k=1}^m \lambda_k \mathcal{B}(x_k,y',u_k,v_k,t_k). \tag{3.6}$$

Here $\lambda_1, \lambda_2, \ldots, \lambda_m \geq 0$ are nonnegative numbers summing up to 1 and the points (x,y,u,v,t), (x_1,y',u_1,v_1,t_1), $\ldots$, (x_m,y',u_m,v_m,t_m) are elements of D enjoying the following conditions: we have $v_j = \max\{u_j, v\}$ for all $j = 1, 2, \ldots, m$ and

$$x = \sum_{k=1}^m \lambda_k x_k, \qquad y' = y + \left(t - \sum_{k=1}^m \lambda_k t_k\right)x, \qquad u = \sum_{k=1}^m \lambda_k u_k, \qquad t \geq \sum_{k=1}^m \lambda_k t_k.$$

Proof. First we apply 3° with $\delta = t - \sum_{k=1}^m \lambda_k t_k$ to get

$$\mathcal{B}(x,y,u,v,t) \geq \mathcal{B}\left(x,y',u,v,\sum_{k=1}^m \lambda_k t_k\right).$$

Therefore, we will be done if we prove that

$$\mathcal{B}\left(x, y', u, v, \sum_{k=1}^{m} \lambda_k t_k\right) \geq \sum_{k=1}^{m} \lambda_k \mathcal{B}(x_k, y', u_k, v_k, t_k).$$

This follows easily by the induction argument. If $m = 2$, then the above bound reduces to 4°. For the inductive step, suppose that u_m is the biggest element of the set $\{u_1, u_2, \ldots, u_m\}$ and set $\lambda = \sum_{k=1}^{m-1} \lambda_k$, $x_- = \lambda^{-1} \sum_{k=1}^{m-1} \lambda_k x_k$, $x_+ = x_m$, $u_- = \lambda^{-1} \sum_{k=1}^{m-1} \lambda_k u_k$, $u_+ = u_m$, $v_\pm = \max\{u_\pm, v\}$, $t_- = \lambda^{-1} \sum_{k=1}^{m-1} \lambda_k t_k$ and $t_+ = t_m$. Since u_m is the biggest of u_j's, we have $u_- \leq u \leq v$ and hence $v_- = v$. Consequently, the condition 4° implies

$$\mathcal{B}\left(x, y', u, v, \sum_{k=1}^{m} \lambda_k t_k\right) \geq \lambda \mathcal{B}(x_-, y', u_-, v, t_-) + \lambda_m \mathcal{B}(x_m, y', u_m, v_m, t_m),$$

and, by the induction hypothesis,

$$\mathcal{B}(x_-, y', u_-, v, t_-) \geq \lambda^{-1} \sum_{k=1}^{m-1} \lambda_k \mathcal{B}(x_k, y', u_k, v_k, t_k).$$

Putting all the facts together, we get the claim. □

In what follows, we say that a function f on X is $\mathcal{T}$-simple, if it is measurable with respect to the σ-algebra generated by $\mathcal{T}^N$ for some integer N.

Theorem 3.2. *If there is a function $\mathcal{B}$ satisfying 1°, 2°, 3° and 4°, then* (3.1) *holds true for any probability space (X, μ) with a tree $\mathcal{T}$, any $\mathcal{T}$-simple function $f : X \to [0, \infty)$, any $\mathcal{T}$-simple weight $w \in A_1$ satisfying $[w]_{A_1} \leq c$ and any* 1-*Carleson sequence α having a finite number of nonzero terms.*

Proof. Fix (X, μ), $\mathcal{T}$ and any f, w and α as in the statement. We split the reasoning into three intermediate parts.

Step 1. Auxiliary notation. For any $n \geq 0$, define the functions f_n, g_n, w_n, w_n^*, σ_n on X as follows: if $\omega \in X$ and $Q = Q_n(\omega)$ denotes the unique element of $\mathcal{T}^n$ which contains ω, then

$$\begin{aligned}
f_n(\omega) &= \langle f \rangle_{Q,\mu}, \\
g_n(\omega) &= \sum_{R \in \mathcal{T},\, R \supsetneq Q} \alpha_R \langle f \rangle_{R,\mu} \chi_R(\omega), \\
w_n(\omega) &= \langle w \rangle_{Q,\mu}, \\
w_n^*(\omega) &= \max_{k \leq n} w_k(\omega) \\
\sigma_n(\omega) &= \frac{1}{\mu(Q)} \sum_{R \in \mathcal{T},\, R \subseteq Q} \alpha_R \mu(R).
\end{aligned}$$

It is easy to see that $(f_n, g_n, w_n, w_n^*, \sigma_n)$ takes values in the set D: this is the consequence of the inequality $[w]_{A_1} \leq c$ and the 1-Carleson property of α.

Step 2. Monotonicity. Now we will prove that

$$\text{the sequence } \left(\int_X \mathcal{B}(f_n, g_n, w_n, w_n^*, \sigma_n)\mathrm{d}\mu\right)_{n\geq 0} \text{ is nonincreasing.} \tag{3.7}$$

This is a simple combination of the inequality (3.6) and the evolution rules of (f, g, w, w^*, σ). Namely, fix $n \geq 0$, an element $Q \in \mathcal{T}^n$ and denote the children of Q in $\mathcal{T}^{n+1}$ by Q_1, Q_2, …, Q_m. The functions f_n, g_n, w_n, w_n^* and σ_n are constant on Q: denote the corresponding values by x, y, u, v and t. Similarly, f_{n+1}, g_{n+1}, w_{n+1}, w_{n+1}^* and σ_{n+1} are constant on each Q_j: denote the values by x_j, y_j, u_j, v_j and t_j, respectively. Actually, since g_{n+1} is $\mathcal{T}^n$ measurable, we see that all y_j's are equal: denote this common value by y'. Let us check that the conditions listed below (3.6) are satisfied, with $\lambda_j = \mu(Q_j)/\mu(Q)$. The numbers λ_j sum up to 1 and

$$x = \frac{1}{\mu(Q)}\int_Q f\mathrm{d}\mu = \sum_{k=1}^m \frac{\mu(Q_k)}{\mu(Q)} \cdot \frac{1}{\mu(Q_k)}\int_{Q_k} f\mathrm{d}\mu = \sum_{k=1}^m \lambda_k x_k.$$

The identities $u = \sum_{k=1}^m \lambda_k u_k$ and $t = \alpha_Q + \sum_{k=1}^m \lambda_k t_k \geq \sum_{k=1}^m \lambda_k t_k$ are verified analogously. Moreover, for each j we obviously have

$$v_j = \max_{k\leq n+1} w_k|_{Q_j} = \max\left\{w_{n+1}|_{Q_j}, \max_{k\leq n} w_k|_{Q_j}\right\} = \max\{u_j, v\}$$

and

$$y' = \sum_{R\in\mathcal{T},\, R\supsetneq Q} \alpha_R\langle f\rangle_{R,\mu} = \sum_{R\in\mathcal{T},\, R\supsetneq Q} \alpha_R\langle f\rangle_{R,\mu} + \alpha_Q\langle f\rangle_{Q,\mu} = y + \left(t - \sum_{k=1}^m \lambda_k t_k\right)x.$$

Consequently, we may apply (3.6), and this estimate is equivalent to

$$\int_Q \mathcal{B}(f_n, g_n, w_n, w_n^*, \sigma_n)\mathrm{d}\mu \geq \int_Q \mathcal{B}(f_{n+1}, g_{n+1}, w_{n+1}, w_{n+1}^*, \sigma_{n+1})\mathrm{d}\mu.$$

Summing over all $Q \in \mathcal{T}^n$, we get the desired monotonicity.

Step 3. Completion of the proof. Fix a large integer N such that f, w are $\sigma(\mathcal{T}^{N+1})$ measurable and such that $\alpha_Q = 0$ for any $Q \in \mathcal{T}^n$, $n \geq N$ (such an N exists due to the simplicity assumptions and the vanishing of almost all terms of α). By the previous step, we get

$$\int_X \mathcal{B}(f_N, g_N, w_N, w_N^*, \sigma_N)\mathrm{d}\mu \leq \int_X \mathcal{B}(f_0, g_0, w_0, w_0^*, \sigma_0)\mathrm{d}\mu.$$

But $g_0 = 0$, so by (3.2), the right-hand side is nonpositive. Furthermore, we have $f_N = N$, $g_N = \mathcal{A}_\alpha^{X,\mathcal{T}} f$, $w_N = w$ and $\sigma_N = 0$, so applying (3.3) to the left-hand side, we get the claim. □

Now we will handle the implication in the reverse direction.

Theorem 3.3. *The reverse to Theorem 3.2 holds true.*

Proof. Introduce the function $\mathcal{B}: D \to \mathbb{R}$ by the formula

$$\mathcal{B}(x,y,u,v,t) = \sup\left\{ \int_X V\Big(f, y + \mathcal{A}_\alpha^{X,\mathcal{T}} f, w\Big) \mathrm{d}\mu \right\}.$$

Here the supremum is taken over all probability spaces X with a tree $\mathcal{T}$, all $\mathcal{T}$-simple functions $f : X \to [0,\infty)$ satisfying $\int_X f \mathrm{d}\mu = x$, all $\mathcal{T}$-simple A_1 weights $w : X \to [c^{-1}v, \infty)$ satisfying $[w]_{A_p} \leq c$, $\int_X w \mathrm{d}\mu = u$ and all 1-Carleson sequences $\alpha = (\alpha_R)_{R\in\mathcal{T}}$ having a finite number of nonzero terms and satisfying $\sum_{R\in\mathcal{T}} \alpha_R \mu(R) \leq t$.

We will now verify that $\mathcal{B}$ enjoys the properties 1°, 2° and 3°. The initial condition follows directly from (3.1): indeed, for any X, $\mathcal{T}$, f, w and α as in the definition of $\mathcal{B}(x,0,u,v,t)$ we have

$$\int_X V\Big(f, \mathcal{A}_\alpha^{X,\mathcal{T}} f, w\Big) \mathrm{d}\mu \leq 0,$$

and the inequality remains valid if we take the supremum. The majorization is also very simple: pick arbitrary X, $\mathcal{T}$ and consider the constant function $f \equiv x$, the constant weight $w \equiv u$ and the Carleson sequence containing only zeros. By the very definition of $\mathcal{B}$, we may write

$$\mathcal{B}(x,y,u,v,0) \geq \int_X V\Big(f, y + \mathcal{A}_\alpha^{X,\mathcal{T}} f, w\Big) \mathrm{d}\mu = V(x,y,u).$$

We take the opportunity to record here a slightly stronger majorization condition, which will be useful later. Namely, if $(x,y,u,v,t) \in D$, then

$$\mathcal{B}(x,y,u,v,t) \geq V(x, y + tx, u). \tag{3.8}$$

This can be seen by taking, in the definition of $\mathcal{B}(x,y,u,v,t)$, the functions f, w as previously and the Carleson sequence containing only zeros, except for the term $\alpha_X = t$.

To show 3°, pick x, y, u, v, t and δ as in the statement and any X, $\mathcal{T}$, f, w and α as in the definition of $\mathcal{B}(x, y + \delta x, u, v, t - \delta)$. It is easy to check that the modified sequence $\alpha' = (\alpha'_R)_{R\in\mathcal{T}}$ given by $\alpha'_X = \alpha_X + \delta$ and $\alpha'_R = \alpha_R$ for all $R \neq X$ has 1-Carleson property, a finite number of nonzero terms and satisfies $\sum_{R\in\mathcal{T}} \alpha'_R \mu(R) \leq t$. Because $\mathcal{A}_{\alpha'}^{X,\mathcal{T}} f = \delta x + \mathcal{A}_\alpha^{X,\mathcal{T}} f$, we may write

$$\mathcal{B}(x,y,u,v,t) \geq \int_X V\Big(f, y + \mathcal{A}_{\alpha'}^{X,\mathcal{T}} f, w\Big) \mathrm{d}\mu = \int_X V\Big(f, y + \delta x + \mathcal{A}_\alpha^{X,\mathcal{T}} f, w\Big) \mathrm{d}\mu,$$

so (3.4) follows, since X, $\mathcal{T}$, f, w and α were arbitrary. It remains to prove the concavity-type condition 4°. Pick the parameters $x_\pm$, y, $u_\pm$, $\ldots$ as in the statement and fix an auxiliary number $\varepsilon > 0$. By the definition of $\mathcal{B}$, there are probability spaces $(X_\pm, \mu_\pm)$ with a tree $\mathcal{T}_\pm$, as well as appropriate functions $f_\pm$, $w_\pm$ and 1-Carleson sequences $\alpha^\pm$, such that

$$\mathcal{B}(x_\pm, y, u_\pm, v_\pm, t_\pm) \leq \int_{X_\pm} V\left(f_\pm, y + \mathcal{A}_{\alpha^\pm}^{X_\pm,\mathcal{T}_\pm} f_\pm, w_\pm\right) \mathrm{d}\mu_\pm + \varepsilon. \tag{3.9}$$

With no loss of generality, we may assume that $X_\pm$ are pairwise disjoint. We splice them into one space $X = X_- \cup X_+$ with the probability measure μ given by $\mu(A) = \lambda\mu_-(A \cap X_-) + (1-\lambda)\mu_+(A \cap X_+)$ and the tree structure $\mathcal{T}$ such that $\mathcal{T}^0 = \{X\}$ and $\mathcal{T}^n = \mathcal{T}_-^{n-1} \cup \mathcal{T}_+^{n-1}$ for $n \geq 1$. Next, we "splice" the functions, weights and Carleson sequences as follows. Define $f, w : X \to [0, \infty)$ by $f = f_-\chi_{X_-} + f_+\chi_{X_+}$ and $w = w_-\chi_{X_-} + w_+\chi_{X_+}$, and set

$$\alpha_R = \begin{cases} 0 & \text{if } R = X, \\ \alpha_R^- & \text{if } R \in \mathcal{T}_-, \\ \alpha_R^+ & \text{if } R \in \mathcal{T}_+. \end{cases}$$

Let us check that f, w and α satisfy all the requirements in the definition of $\mathcal{B}(x, y, u, v, t)$. First, note that

$$\int_X f \mathrm{d}\mu = \int_{X_-} f_- \mathrm{d}(\lambda\mu_-) + \int_{X_+} f_+ \mathrm{d}((1-\lambda)\mu_+) = \lambda x_- + (1-\lambda)x_+ = x$$

and similarly, $\int_X w \mathrm{d}x = u$. In addition, since $w_\pm$ takes values in $[c^{-1}v_\pm, \infty)$ and $v_\pm \geq v$, we see that $w \in [c^{-1}v, \infty)$. Furthermore, for any $Q \in \mathcal{T}_\pm$,

$$\langle w \rangle_{Q,\mu} \geq c^{-1}v_\pm \geq c^{-1}v \geq c^{-1}u = c^{-1}\langle w \rangle_{X,\mu}$$

and, since $w_\pm$ is an A_1 weight satisfying $[w_\pm]_{A_1} \leq c$,

$$\langle w \rangle_{Q,\mu} = \langle w_\pm \rangle_{Q,\mu_\pm} \geq c^{-1} \max_{R \supseteq Q, Q \in \mathcal{T}_\pm} \langle w_\pm \rangle_{R,\mu_\pm} = c^{-1} \max_{R \supseteq Q, Q \in \mathcal{T}_\pm} \langle w \rangle_{R,\mu}.$$

Combining the two bounds above, we conclude that $[w]_{A_1} \leq c$. It remains to check that α is a 1-Carleson sequence satisfying $\sum_{Q \in \mathcal{T}} \alpha_Q \mu(Q) \leq t$. This is very simple: by the very definition of α and the properties of $\alpha^\pm$,

$$\begin{aligned} \sum_{Q \in \mathcal{T}} \alpha_Q \mu(Q) &= \alpha_X \mu(X) + \lambda \sum_{Q \in \mathcal{T}_-} \alpha_Q^- \mu_-(Q) + (1-\lambda) \sum_{Q \in \mathcal{T}_+} \alpha_Q^+ \mu_+(Q) \\ &\leq \lambda t_- + (1-\lambda)t_+ = t. \end{aligned}$$

To verify the 1-Carleson property, note that if $Q \in \mathcal{T}_\pm$, then

$$\frac{1}{\mu(Q)} \sum_{R \supseteq Q, R \in \mathcal{T}} \alpha_R \mu(R) = \frac{1}{\mu_\pm(Q)} \sum_{R \supseteq Q, R \in \mathcal{T}_j} \alpha_R \mu_\pm(R) \leq 1,$$

while for $Q = X$ we have just checked that $\mu(Q)^{-1} \sum_{Q \in \mathcal{T}} \alpha_Q \mu(Q) \leq t \leq 1$. Thus f, w and α have all the required properties and hence

$$\mathcal{B}(x, y, u, v, t) \geq \int_X V\left(f, y + \mathcal{A}_\alpha^{X,\mathcal{T}} f, w\right) \mathrm{d}\mu.$$

Since $\alpha_X = 0$, we have $\mathcal{A}_\alpha^{X,\mathcal{T}} f = \mathcal{A}_{\alpha^-}^{X_-,\mathcal{T}_-} f_- \chi_{X_-} + \mathcal{A}_{\alpha^-}^{X_-,\mathcal{T}_-} f_- \chi_{X_+}$ and therefore the latter integral equals

$$\int_{X_-} V\Big(f_-, y + \mathcal{A}_{\alpha^-}^{X_-,\mathcal{T}_-} f_-, w_-\Big)\mathrm{d}\mu + \int_{X_+} V\Big(f_+, y + \mathcal{A}_{\alpha^+}^{X_+,\mathcal{T}_+} f_+, w_+\Big)\mathrm{d}\mu$$
$$\geq \lambda\mathcal{B}(x_-, y, u_-, v_-, t_-) + (1-\lambda)\mathcal{B}(x_+, y, u_+, v_+, t_+) - \varepsilon,$$

where in the last passage we have exploited (3.9). Since ε was arbitrary, the concavity condition follows. □

4. Dimension-free estimates for shift operators

We start with an auxiliary weighted L^2 estimate, which will be proved directly. We will use the standard notation $\langle f\rangle_{Q,w} = \frac{1}{w(Q)}\int_Q fw\mathrm{d}\mu$ for the average of f over Q with respect to the measure w. Furthermore, M_w will stand for the associated maximal function: $M_w f = \sup_Q \Big\{\langle f\rangle_{Q,w}\chi_Q\Big\}$.

Theorem 4.1. *Suppose that w is an A_1 weight on X and α is a 1-Carleson sequence. Then we have $\|\mathcal{A}_\alpha^{X,\mathcal{T}}\|_{L^2(w^{-1})\to L^2(w^{-1})} \leq 4[w]_{A_1}$.*

Proof. For any $f \in L^2(w^{-1})$ we have, by duality,

$$\|\mathcal{A}_\alpha^{X,\mathcal{T}} f\|_{L^2(w^{-1})} = \sup\left\{\int_X g\mathcal{A}_\alpha^{X,\mathcal{T}} f\mathrm{d}\mu \,:\, \|g\|_{L^2(w)} = 1\right\}.$$

Let $(E(Q))_{Q\in\mathcal{T}}$ be a pairwise disjoint family of subsets of X associated with a given 1-Carleson sequence α. For any $g \in L^2(w)$ we have

$$\begin{aligned}
&\int_X g\mathcal{A}_\alpha^{X,\mathcal{T}} f\mathrm{d}\mu \\
&\quad = \sum_{Q\in\mathcal{T}} \mu(Q)\langle f\rangle_{Q,\mu}\langle g\rangle_{Q,\mu} \\
&\quad = \sum_{Q\in\mathcal{T}} \mu(E(Q))\langle w\rangle_{Q,\mu}\langle w^{-1}\rangle_{Q,\mu}\langle fw^{-1}\rangle_{Q,w}\langle gw\rangle_{Q,w^{-1}} \\
&\quad \leq [w]_{A_1} \sum_{Q\in\mathcal{T}} w(E(Q))^{1/2}\langle fw^{-1}\rangle_{Q,w} \cdot \big(w^{-1}(E(Q))\big)^{1/2}\langle gw\rangle_{Q,w^{-1}} \\
&\quad \leq [w]_{A_1}\left(\sum_{Q\in\mathcal{T}} \langle fw^{-1}\rangle_{Q,w}^2 w(E(Q))\right)^{1/2}\left(\sum_{Q\in\mathcal{T}} \langle gw\rangle_{Q,w^{-1}}^2 w^{-1}(E(Q))\right)^{1/2} \\
&\quad \leq [w]_{A_1}\|M_w(fw^{-1})\|_{L^2(w)}\|M_w(gw)\|_{L^2(w^{-1})} \\
&\quad \leq 4[w]_{A_1}\|f\|_{L^2(w^{-1})}\|g\|_{L^2(w)},
\end{aligned}$$

where in the last line we have used the unweighted L^2 bound for the maximal function. This gives the claim. □

Now we apply the Bellman function method. The theorem above implies that if we put $V(x,y,u)=y^2u^{-1}-16c^2x^2u^{-1}$, then we have

$$\int_X V(f,\mathcal{A}_\alpha^{X,\mathcal{T}}f,w)\mathrm{d}\mu\leq 0,$$

for any probability space (X,μ) with a tree $\mathcal{T}$, any $\mathcal{T}$-simple function $f:X\to[0,\infty)$, any $\mathcal{T}$-simple weight $w\in A_1$ satisfying $[w]_{A_1}\leq c$ and any 1-Carleson sequence α having a finite number of nonzero terms. Therefore, by Theorem 3.3 there exists an associated function $\mathcal{B}$ possessing the properties 1°, 2°, 3° and 4°. Now we will modify $\mathcal{B}$ to obtain the Bellman corresponding to the dual estimate of Muckehoupt and Wheeden. Define $\overline{\mathcal{B}}$, $\overline{V}:D\to\mathbb{R}$ by

$$\overline{\mathcal{B}}(x,y,u,v,t)=\begin{cases}\min\Big\{\mathcal{B}(x,y,u,v,t),u-8c|x|\Big\} & \text{if } |x|<u/(4c),\\ u-8c|x| & \text{if } |x|\geq u/(4c)\end{cases}\tag{4.1}$$

and $\overline{V}(x,y,u,v,t)=u\chi_{\{y\geq v\}}-8c|x|$. Obviously, we have

$$\overline{\mathcal{B}}(x,y,u,v,t)\leq u-8c|x|\qquad \text{on } D.\tag{4.2}$$

Furthermore, by (3.8), if $|x|=u/(4c)$, we have

$$\mathcal{B}(x,y,u,v,t)\geq\left(y+\frac{tu}{4c}\right)^2u^{-1}-u\geq -u=u-8c|x|,$$

so we also have

$$\overline{\mathcal{B}}(x,y,u,v,t)=\begin{cases}\min\Big\{\mathcal{B}(x,y,u,v,t),u-8c|x|\Big\} & \text{if } |x|\leq u/(4c),\\ u-8c|x| & \text{if } |x|> u/(4c).\end{cases}$$

(in comparison to the formula (4.1), the inequalities $|x|<u/(4c)$ and $|x|\geq u/(4c)$ have become non-strict and strict, respectively). We will prove the following fact.

Theorem 4.2. *The function $\overline{\mathcal{B}}$ satisfies the conditions 1°, 2°, 3° and 4°, with respect to $\overline{V}$.*

Proof. The condition 1°. This is easy to check: if $|x|\leq u/(4c)$, then we have $\overline{\mathcal{B}}(x,0,u,v,t)\leq\mathcal{B}(x,y,u,v,t)\leq 0$, by the initial property of $\mathcal{B}$; on the other hand, if $|x|>u/(4c)$, then $\overline{\mathcal{B}}(x,0,u,v,t)=-8c|x|\leq 0$.

The condition 2°. Suppose first that $y\geq v$. Then the majorization is trivial, since $\overline{\mathcal{B}}(x,y,u,v,t)=u-8c|x|$, by the very definition of $\overline{\mathcal{B}}$. Indeed, we only need to check that for $|x|<u/(4c)$ the minimum is attained at the expression $u-8c|x|$: applying (3.8), we get

$$\mathcal{B}(x,y,u,v,t)\geq(y+tx)^2u^{-1}-16c^2x^2u^{-1}\geq v^2u^{-1}-16c^2x^2u^{-1}\geq u-8c|x|.$$

So, it remains to verify 2° for $y<v$; then the desired bound becomes

$$\overline{\mathcal{B}}(x,y,u,v,t)\geq -8c|x|.$$

This is obvious if $\overline{\mathcal{B}}(x,y,u,v,t)=u-8c|x|$; otherwise, we must have $|x|<u/(4c)$ and again by (3.8),

$$\begin{aligned}\overline{\mathcal{B}}(x,y,u,v,t)=\mathcal{B}(x,y,u,v,t)\geq (y+tx)^2u^{-1}-16c^2x^2u^{-1}&\geq -16c^2x^2u^{-1}\\ &\geq -8c|x|.\end{aligned}$$

The monotonicity condition 3°. If $\overline{\mathcal{B}}(x,y,u,v,t)=u-8c|x|$, then the inequality follows at once from (4.2): $\overline{\mathcal{B}}(x,y+\delta x,u,v,t-\delta)\leq u-8c|x|$. The remaining case is $\overline{\mathcal{B}}(x,y,u,v,t)=\mathcal{B}(x,y,u,v,t)<u-8c|x|$, and applying the monotonicity property of $\mathcal{B}$ we get

$$\mathcal{B}(x,y,u,v,t)\geq \mathcal{B}(x,y+\delta x,u,v,t-\delta)\geq \overline{\mathcal{B}}(x,y+\delta x,u,v,t-\delta)$$

(the last inequality follows from $|x|<u/(4c)$, which holds since $\overline{\mathcal{B}}(x,y,u,v,t)<u-8c|x|$).

The concavity-type property 4°. If $\overline{\mathcal{B}}(x,y,u,v,t)=u-8c|x|$, then the condition follows directly from (4.2) and the convexity of the function $r\mapsto |r|$:

$$\begin{aligned}\overline{\mathcal{B}}\,(x,y,u,v,t)=u-8c|x|&\geq \lambda(u_--8c|x_-|)+(1-\lambda)(u_+-8c|x_+|)\\ &\geq \lambda\overline{\mathcal{B}}(x_-,y,u_-,v_-,t_-)+(1-\lambda)\overline{\mathcal{B}}(x_+,y,u_+,v_+,t_+).\end{aligned}$$

So, suppose that $\overline{\mathcal{B}}(x,y,u,v,t)=\mathcal{B}(x,y,u,v,t)<u-8c|x|$ (hence, in particular, $|x|<u/(4c)$). If we have $|x_\pm|\leq u_\pm/(4c)$, then the claim follows from the definition of $\overline{\mathcal{B}}$ and the concavity-type condition for $\mathcal{B}$:

$$\begin{aligned}\overline{\mathcal{B}}\,(x,y,u,v,t)&=\mathcal{B}\,(x,y,u,v,t)\\ &\geq \lambda\mathcal{B}\,(x_-,y,u_-,v_-,t_-)+(1-\lambda)\mathcal{B}\,(x_+,y,u_+,v_+,t_+)\\ &\geq \lambda\overline{\mathcal{B}}(x_-,y,u_-,v_-,t_-)+(1-\lambda)\overline{\mathcal{B}}(x_+,y,u_+,v_+,t_+).\end{aligned}$$

The case in which $|x_-|>u_-/(4c)$ or $|x_+|>u_+/(4c)$ follows from the above inequality, the convexity of the function $r\mapsto|r|$ and a 'truncation argument'. For instance, suppose that $|x_-|>u_-/(4c)$. Then there is a number $s\in(0,1)$ such that $|x+s(x_--x)|=(u+s(u_--u))/(4c)$: this follows by Darboux property and the fact that for $s\in\{0,1\}$ we have inequalities in reverse directions). Set $x_-^*=x+s(x_--x)$, $u^*=u+s(u_--u)$, $v_-^*=\max\{u_-^*,v\}$ and $t_-^*=t+s(t_--t)$. If $|x_+|>u_+/(4c)$, define x_+^*, u_+^*, v_+^* and t_+^* analogously (if $|x_+|\leq u_+/(4c)$, set $x_+^*=x_+$, $u_+^*=u_+$, $v_+^*=v_+$ and $t_+^*=t_+$). Then $|x_\pm^*|\leq u_\pm^*/(4c)$, so we may write

$$\begin{aligned}\overline{\mathcal{B}}\,(x,y,u,v,t)&=\mathcal{B}\,(x,y,u,v,t)\\ &\geq \lambda^*\overline{\mathcal{B}}(x_-^*,y,u_-^*,v_-^*,t_-^*)+(1-\lambda^*)\overline{\mathcal{B}}(x_+^*,y,u_+^*,v_+^*,t_+^*)\end{aligned}\tag{4.3}$$

for an appropriate $\lambda^*\in(0,1)$ (determined by the condition $\lambda^*x_-^*+(1-\lambda^*)x_+^*=x$). This can be viewed as a truncated version of (3.5). To pass from $(x_-^*,y,u_-^*,v_-^*,t_-^*)$ to (x_-,y,u_-,v_-,t_-), we apply the aforementioned convexity of $r\mapsto|r|$ to get

$$\begin{aligned}\overline{\mathcal{B}}(x_-^*,y,u_-^*,v_-^*,t_-^*)&=u_-^*-8c|x_-^*|\\ &\geq \theta(u-8c|x|)+(1-\theta)(u_--8c|x_-|)\\ &\geq \theta\overline{\mathcal{B}}(x,y,u,v,t)+(1-\theta)\overline{\mathcal{B}}(x_-,y,u_-,v_-,t_-),\end{aligned}$$

where in the last line we exploited (4.2) (and $\theta \in (0,1)$ is determined by the condition $x_-^* = \theta x + (1-\theta)x_-$). We apply the same procedure to $\overline{\mathcal{B}}(x_+^*, y, u_+^*, v_+^*, t_+^*)$ if $|x_+| > u_+/(4c)$ (if $|x_+| \leq u_+/(4c)$, then we just write $\overline{\mathcal{B}}(x_+^*, y, u_+^*, v_+^*, t_+^*) = \overline{\mathcal{B}}(x_+, y, u_+, v_+, t_+)$). Combining these facts with (4.3), we obtain the desired concavity. □

The existence of the Bellman function immediately yields the weighted estimate for shift operators.

Proof of (1.7). Let (X, μ) be a probability space with a tree structure $\mathcal{T}$ and assume that $f : X \to [0, \infty)$ is an arbitrary integrable function, w is an A_1 weight on X (with $c := [w]_{A_1}$) and $\alpha = (\alpha_Q)_{Q \in \mathcal{T}}$ is an η-Carleson sequence. Replacing α with α/η, we may assume that $\eta = 1$. We approximate f, w and α by appropriate simple objects: given a large positive integer N, we let f_N, w_N be the conditional expectations of f and w with respect to $\mathcal{T}^N$ (see the proof of Theorem 3.2, Step 1), and introduce the truncated sequence $\alpha^N = (\alpha_Q')_{Q \in \mathcal{T}}$ by $\alpha_Q^N = \alpha_Q$ if $Q \in \mathcal{T}^n$ for some $n \leq N$, and $\alpha_Q^N = 0$ otherwise. Then f, w are $\mathcal{T}$-simple and α' is a 1-Carleson sequence containing only a finite number of nonzero terms. Therefore, by Theorem 3.2 applied to $\overline{\mathcal{B}}$ and $\overline{V}$ we get

$$\int_X w_N(\mathcal{A}_{\alpha^N} f_N \geq w_N) \mathrm{d}\mu \leq 8[w]_{A_1} \int_X f_N \mathrm{d}\mu = 8[w]_{A_1} \int_X f \mathrm{d}\mu.$$

Note that $\int_X w_N(\mathcal{A}_{\alpha^N} f_N \geq w_N)\mathrm{d}\mu = \int_X w(\mathcal{A}_{\alpha'} f_N \geq w_N)\mathrm{d}\mu$. In addition, we have $\mathcal{A}_{\alpha^N} f_N \to \mathcal{A}_\alpha f$ and $w_N \to w$ μ-almost surely as $N \to \infty$, so the previous estimate yields

$$\int_X w(\mathcal{A}_\alpha f > w)\mathrm{d}\mu \leq 8[w]_{A_1} \int_X f \mathrm{d}\mu.$$

To obtain the non-strict inequality on the left, consider an auxiliary parameter $\theta \in (0,1)$ and apply the above bound to the A_1 weight θw:

$$\int_X w(\mathcal{A}_\alpha f \geq w)\mathrm{d}\mu \leq \int_X w(\mathcal{A}_\alpha f > \theta w)\mathrm{d}\mu \leq 8\theta^{-1}[\theta w]_{A_1} \int_X f \mathrm{d}\mu.$$

Since $[\theta w]_{A_1} = [w]_{A_1}$, letting $\theta \to 1$ completes the proof. □

5. Weighted inequality for singular integrals

We begin by recalling some basic information about the class of Calderón-Zygmund singular integral operators to be studied below. We consider kernels K defined on $\mathbb{R}^d \times \mathbb{R}^d \setminus \{(x,x) : x \in \mathbb{R}^d\}$ which satisfy the following two requirements:

(i) The shift-invariant size estimate: for some $C > 0$ we have

$$|K(x,y)| \leq \frac{C}{|x-y|^d} \quad \text{for all } x \neq y.$$

(ii) Continuity of the kernel off the diagonal: we have

$$|K(x,y)-K(x',y)| \le \frac{|x-x'|^\delta}{(|x-y|+|x'-y|)^{n+\delta}}$$

if $|x-x'| \le \frac{1}{2}\max\{|x-y|,|x'-y|\}$, and

$$|K(x,y)-K(x,y')| \le \frac{|y-y'|^\delta}{(|x-y|+|x-y'|)^{n+\delta}}$$

whenever $|y-y'| \le \frac{1}{2}\max\{|x-y|,|x-y'|\}$.

Consider a continuous linear operator T from the space $\mathscr{S}(\mathbb{R}^d)$ of Schwartz functions to the space $\mathscr{S}'(\mathbb{R}^d)$ of all tempered distributions. Then T is associated with K, if it satisfies

$$T(f)(x) = \int_{\mathbb{R}^d} K(x,y)f(y)\mathrm{d}y$$

for all $f \in \mathscr{C}_0^\infty$ and x not in the support of f. If T is associated with K and admits the bounded extension on $L^2(\mathbb{R}^d)$, then T will be referred to as the Calderón-Zygmund operator.

The above class of singular integral operators on $\mathbb{R}^d$ is controlled by the family of dyadic shifts on $\mathbb{R}^d$ (with respect to general, not necessarily standard) dyadic grids in the sense specified below. Recall the following fact, which is the consequence of the results proved by Lerner and Nazarov (see Theorem 13.2 in [10]).

Theorem 5.1. *Let T be a Calderón-Zygmund operator on $\mathbb{R}^d$, associated with a kernel K. Then for any integrable function $f:\mathbb{R}^d\to\mathbb{R}$ there exist 3^d dyadic lattices $\mathscr{D}^{(j)}$ and $14\cdot 3^d$-Carleson families $\mathcal{S}_j \subset \mathscr{D}^{(j)}$, $j=1,2,\ldots,3^d$, such that*

$$|Tf| \le C_{T,d}\sum_{j=1}^{3^d} \mathcal{A}_{\mathcal{S}_j}|f|$$

almost everywhere. The constant $C_{T,d}$ depends on the parameters indicated.

Now it is easy to establish the weak dual conjecture of Muckenhoupt and Wheeden for Calderón-Zygmund operators.

Proof of (1.4). We combine Theorem 5.1 with the weak-type bound (1.7) for dyadic shifts. As the result we get, for any f,

$$\begin{aligned} w(|Tf|\ge w) &\le w\left(C_{T,d}\sum_{j=1}^{3^d}\mathcal{A}_{\mathcal{S}_j}|f| \ge w\right)\\ &\le \sum_{j=1}^{3^d} w(C_{T,d}\cdot 3^d\mathcal{A}_{\mathcal{S}_j}|f|\ge w)\\ &\le C_{T,d}\cdot 3^d\cdot 14\cdot 3^d[w]_{A_1}\int_{\mathbb{R}^d}|f|\mathrm{d}x \end{aligned}$$

and we are done. □

On the linear dependence on the characteristic. Let H be the Hilbert transform, the singular integral operator acting on integrable functions f by

$$Hf(x) = \frac{1}{\pi}\text{p.v.}\int_{\mathbb{R}} \frac{f(y)}{x-y}\mathrm{d}y.$$

Suppose that the weak-type estimate

$$w(|Hf| \geq w) \leq C[w]_{A_1}^{\kappa}\|f\|_{L^1(\mathbb{R})} \tag{5.1}$$

holds for arbitrary integrable function f and any A_1 weight w, where C, κ are finite universal constants (not depending on f or w). Let $\gamma \in (0,1)$ be a fixed parameter and set $\beta = 2^{(2-\gamma)/(\gamma-1)}$ (so that $(2\beta)^{1-\alpha} = 1/2$). It is well known that the power weight $w(x) = \pi^{-1}|x+\beta|^{-\gamma}$ satisfies the condition A_1 with $(1-\gamma)^{-1} \leq [w]_{A_1} \leq 2(1-\gamma)^{-1}$. Furthermore, setting $f = (2\beta)^{-1}\chi_{[-\beta,\beta]}$, we have $\int_{\mathbb{R}} f = 1$ and

$$|Hf(x)| = \frac{1}{2\beta\pi}\left|\log\left|\frac{x+\beta}{x-\beta}\right|\right|$$

for almost all $x \in \mathbb{R}$. Therefore, if $x \in (\beta, 1-\beta)$, then the elementary estimate $\log(1+s) \geq s/(1+s)$ yields

$$|Hf(x)| = \frac{1}{2\beta\pi}\log\left(\frac{x+\beta}{x-\beta}\right) \geq \frac{1}{\pi(x+\beta)} \geq w(x).$$

Consequently, coming back to (5.1), we see that

$$\begin{aligned} C \geq \frac{w(|Hf| \geq w)}{[w]_{A_1}^{\kappa}\int_{\mathbb{R}} f\mathrm{d}x} \geq \frac{(1-\gamma)^{\kappa}}{2^{\kappa}}\int_{\beta}^{1-\beta} w\mathrm{d}x &= \frac{(1-\gamma)^{\kappa-1}}{2^{\kappa}}\frac{1-(2\beta)^{1-\gamma}}{\pi} \\ &= \frac{(1-\gamma)^{\kappa-1}}{2^{\kappa+1}\pi}. \end{aligned}$$

Letting $\gamma \to 1$ we see that κ must be at least one, and hence the linear dependence on the A_1 characteristic of w in (1.4) is the best possible. Obviously, this also implies the same optimality in (1.6) (otherwise, the estimate (1.4) could be improved). □

6. Appendix: an explicit counterexample to (1.3)

Finally, we will construct appropriate functions and weights on $\mathbb{R}$ for which (1.3) will fail to hold. Actually, we will prove a slightly stronger statement involving the Φ-maximal operator. We need some definitions. Suppose that $\Phi : [0,\infty) \to [0,\infty)$ is a Young function, i.e., a continuous, convex and increasing function satisfying $\Phi(0) = 0$ and $\Phi(t) \to \infty$ as $t \to \infty$. The associated maximal operator M_{Φ} acts on locally integrable functions on $\mathbb{R}^d$ by

$$M_{\Phi}f(x) = \sup \|f\|_{\Phi,Q},$$

where the supremum is taken over all cubes $Q \subset \mathbb{R}^d$ containing x, and

$$\|f\|_{\Phi,Q} = \inf\left\{\lambda > 0 \,:\, \frac{1}{|Q|}\int_Q \Phi\left(\frac{|f(y)|}{\lambda}\right)\mathrm{d}y \le 1\right\}.$$

We will show that if the function Φ satisfies

$$\lim_{t\to\infty} \frac{\Phi(t)}{t\log\log t} = 0, \tag{6.1}$$

then the estimate

$$w(\{x \in \mathbb{R}^d : |Tf(x)| > M_\Phi w(x)\}) \le C_{T,d}\int_{\mathbb{R}^d} |f|\mathrm{d}x \tag{6.2}$$

does not hold with any finite constant. This fact has already been proved in [4], but we take the opportunity to present here a different construction. Note that $\|f\|_{\Phi,Q} \le \langle|f|\rangle_Q$ and hence M_Φ is smaller than M, the Hardy-Littlewood maximal operator. So, the above statement gives that the estimate (1.3) fails to hold as well. There is an interesting question whether (6.2) fails for some Young functions outside the class (6.1), e.g., for $\Phi(t) = t\log\log(t+e^e)$; unfortunately, the construction presented below does not seem to work in this case.

Actually, we will assume that $d = 1$ and $T = \mathcal{H}$, the Hilbert transform. The special objects arising in our construction will have a certain fractal-type property, and we start with the description of the corresponding building block. Let R be a positive even integer and consider the function $\varphi^R : \mathbb{R} \to [-1, 1]$ given by

$$\begin{aligned}\varphi^R &= -\sum_{m=0}^{R/2-1} \chi_{[2mR^{-1},(2m+1)R^{-1})} - \chi_{[(2m+1)R^{-1},(2m+2)R^{-1})}\\ &= -\chi_{[0,R^{-1})} + \chi_{[R^{-1},2R^{-1})} - \chi_{[2R^{-1},3R^{-1})} + \ldots + \chi_{[1-R^{-1},1)}.\end{aligned}$$

Obviously, φ^R has integral zero. Furthermore, the function has the following cancellation property: for any interval $J \subset \mathbb{R}$,

$$\int_J \pm\varphi^R \mathrm{d}x \le R^{-1}. \tag{6.3}$$

The properties of $H\varphi^R$, the Hilbert transform of φ^R, are gathered in a separate statement below.

Lemma 6.1. *(i) We have* $H\varphi^R > 0$ *on* $(-\infty, 0) \cup (1, \infty)$.

(ii) Let $\varepsilon, \delta > 0$ *be fixed parameters. Then we have* $|H\varphi^R| \le \delta$ *outside* $[-\varepsilon, 1+\varepsilon]$ *for sufficiently large* R.

(iii) For any $K > 0$ *we have*

$$|\{x \in [0, 1) : H\varphi^R > K\}| \ge 1 - (1 + e^{-\pi K})^{-1/2}.$$

Proof. We compute directly that

$$H\varphi^R(x) = -\frac{1}{\pi}\log\prod_{m=0}^{R/2-1}\frac{\big|(x-2mR^{-1})(x-(2m+2)R^{-1})\big|}{|x-(2m+1)R^{-1}|^2}.$$

Now, if $x\in(-\infty,0)\cup(1,\infty)$, then for each m,

$$\frac{\big|(x-2mR^{-1})(x-(2m+2)R^{-1})\big|}{|x-(2m+1)R^{-1}|^2} = \frac{(x-2mR^{-1})(x-(2m+2)R^{-1})}{(x-(2m+1)R^{-1})^2} < 1,$$

so the product under the logarithm is less than 1 and (i) follows. Therefore, to show (ii), we must prove the estimate $H\varphi^R<\delta$ only. The above calculation gives that for $x\in(-\infty,-\varepsilon)\cup(1+\varepsilon,\infty)$ we have

$$\frac{\big|(x-2mR^{-1})(x-(2m+2)R^{-1})\big|}{|x-(2m+1)R^{-1}|^2} = 1-\frac{1}{R^2(x-(2m+1)R^{-1})^2} \geq 1-\frac{1}{R^2\varepsilon^2}$$

and hence

$$H\varphi^R(x)\leq -\frac{1}{\pi}\log\left(1-\frac{1}{R^2\varepsilon^2}\right)^{R/2}.$$

As $R\to\infty$, the right hand side converges to 0 and hence the uniform boundedness is proved. To establish (iii), we write

$$H\varphi^R(x) = \sum_{m=0}^{R/2-1} -\frac{1}{\pi}\log\frac{\big|(x-2mR^{-1})(x-(2m+2)R^{-1})\big|}{|x-(2m+1)R^{-1}|^2}$$

and observe that each summand is equal to $H\varphi^2$, up to affine mapping which sends $[0,1]$ onto $[2mR^{-1},(2m+2)R^{-1}]$. Therefore in particular, by the property (i) already proved above, for each m we have

$$H\varphi^R(x)\geq -\frac{1}{\pi}\log\frac{\big|(x-2mR^{-1})(x-(2m+2)R^{-1})\big|}{|x-(2m+1)R^{-1}|^2}$$

on the interval $(2mR^{-1},(2m+2)R^{-1})$. Let us restrict ourselves to this interval and check on which subset the right-hand side is bigger than K. Repeating the above calculations, this is equivalent to $\big|1-R^{-2}(x-(2m+1)R^{-1})^{-2}\big|<e^{-\pi K}$, and hence occurs on the set

$$\left(2mR^{-1},2mR^{-1}+\kappa R^{-1}\right)\cup\left((2m+2)R^{-1}-\kappa R^{-1},2mR^{-1}\right),$$

where $\kappa=1-(1+e^{-\pi K})^{-1/2}$. This set has the Lebesgue measure $2\kappa R^{-1}$ and therefore summing over m we get the desired bound. □

We will need the version of the function φ^R on any interval $I\subset\mathbb{R}$: denoting by ℓ_I the affine transformation of I onto $[0,1]$, we set $\varphi^{R,I}=\varphi^R\circ\ell_I$. By the commuting properties of the Hilbert transform, the above lemma holds true for $H\varphi^{R,I}$ as well, with obvious modifications.

We are ready for the construction of the counterexample, which rests on the following inductive procedure. Let $\alpha > 0$, $\varepsilon > 0$ be fixed parameters. By (6.1), if L is a sufficiently large integer, then we have

$$\Phi\left(\frac{2^L}{2^{k+1}\log L^{\alpha}}\right) \le \frac{2^L}{2^{k+1}}, \qquad k = 0, 1, 2, \ldots, L. \tag{6.4}$$

Fix such an integer L and set $K = \log L^{1/\pi}$.

Initial step. Take $w_0 = \chi_{A_0}$, where $A_0 = [0,1]$ is the unit interval. Then we have $Hw_0(x) = \frac{1}{\pi}\log|x/(x-1)|$ and this function is monotone on $(0,1)$. Furthermore, by a simple calculation, if $x \in B_0 := (0, (1+e^{K\pi})^{-1}) \cup ((1+e^{-K\pi})^{-1}, 1)$, then we have $|Hw_0| > K$.

Induction step. Let us write down precisely the assumptions. Suppose that we have constructed a weight w_n on $\mathbb{R}$, a set A_n on which w_n will be modified and a set $B_n \subseteq A_n$ on which $|Hw_n| > 2^n K$. We assume further that $w_n \equiv 2^n$ on A_n and that both A_n and B_n are finite unions of pairwise disjoint subintervals of $[0,1]$. Let us suppose in addition that the interval $(0,1)$ can be split into the finite union of pairwise disjoint intervals such that the function Hw_n is monotone in the interior of each such interval.

Now we describe the induction step. Consider the 'bad' set $A_n \setminus B_n$ (on which the Hilbert transform $|Hw_n|$ is small). It is again a finite union of pairwise disjoint intervals (and perhaps points, but we discard them): $A_n \setminus B_n = I_1 \cup I_2 \cup \ldots \cup I_N$. Splitting these intervals if necessary, we may and do assume that on each I_j the function Hw_n is of constant sign; this follows from the monotonicity property of Hw_n described above. Now, pick an interval I_j for an arbitrary j and let J_j be the subinterval of I_j which has the same center and such that $|I_j \setminus J_j| < \varepsilon/(2^n \cdot N)$. On the set J_j we modify w_j by adding the term $2^n\varphi^{R,J_j}$ or $-2^n\varphi^{R,J^j}$ (where $R = R_{n,j}$ is a large integer, see below) if $Hw_n \ge 0$ or $Hw_n \le 0$ on I_j, respectively. Since $w_n \equiv 2^n$ on A_n, the obtained function w_{n+1} is again nonnegative; actually, on $J_1 \cup J_2 \cup \ldots \cup J_N$ we have $w_{n+1} \in \{0, 2^{n+1}\}$. Put

$$A_{n+1} = \left\{x \in J_1 \cup J_2 \cup \ldots \cup J_N \,:\, w_{n+1} = 2^{n+1}\right\}$$

and $B_{n+1} = \{x \in A_{n+1} \,:\, |Hw_{n+1}(x)| > 2^{n+1}K\}$.

Analysis of the induction step. We easily see that A_{n+1} and B_{n+1} have all the properties listed at the beginning of the previous step. Furthermore, Hw_{n+1} has the required monotonicity, since by the construction, w_{n+1} is a linear combination of characteristic functions of intervals. Since φ^R has integral zero, we have the equality $\int_0^1 w_n = \int_0^1 w_{n+1}$. Next, observe that the modification of w occurs on the sets J_1, J_2, ..., J_N and, by Lemma 6.1 (ii), the Hilbert transform $H(\pm 2^n\varphi^{R,J_j})$ can be made arbitrarily small outside $I_1 \cup I_2 \cup \ldots \cup I_N$ by picking sufficiently large R. In particular, we may assume that $|Hw_{n+1}|$ is still bigger than $2^k K$ on B_k, $k = 0, 1, 2, \ldots, n$.

We repeat the above procedure L times and set $w = w_L$, $f = 2\pi\alpha w_L$. Then we have the identity

$$\int_{\mathbb{R}} f\mathrm{d}x = 2\pi\alpha \int_0^1 w_L \mathrm{d}x = 2\pi\alpha. \tag{6.5}$$

Let us study the behavior of $M_\Phi w$. For any k, if $B_k = I_1 \cup I_2 \cup \ldots \cup I_N$ is the splitting of B_k into the union of pairwise disjoint intervals, then $B_k^o = I_1^o \cup I_2^o \cup \ldots \cup I_N^o$, where I_j^o is the interval having the same center as I_j and satisfying $|I_j^o| = |I_j|/2$. Fix k, pick $u \in B_k^o$ and an arbitrary interval J containing u. If $\sup_j R_{n,j}$ grows sufficiently fast, then (6.3) and the estimate $\|w_k\|_\infty \leq 2^k$ imply that

$$\frac{1}{|J|}\int_J w\mathrm{d}x \leq \frac{1}{|J|}\int_J w_k \mathrm{d}x + 2^k \leq 2^{k+1}.$$

(This follows from the fact that $u \in B_k^o$ is away from the boundary of B_k). But w takes values in the interval $[0, 2^L]$, so the above estimate implies

$$\frac{1}{|J|}\int_J \Phi\left(w(x)/\lambda\right)\mathrm{d}x \leq \frac{2^{k+1}}{2^L}\Phi(2^L/\lambda)$$

for any $\lambda > 0$. Setting $\lambda = 2^{k+1}\log L^\alpha$ and applying (6.4), we get the estimate $\frac{1}{|J|}\int_J \Phi\left(w(x)/\lambda\right)\mathrm{d}x \leq 1$. Since J was chosen arbitrarily, we obtain $M_\Phi w \leq 2^{k+1}\log L^\alpha = 2\pi\alpha \cdot 2^k K < 2\pi\alpha|\mathcal{H}w| = |\mathcal{H}f|$ on B_k^o. Putting all the above facts together, we get

$$\begin{aligned}
w(|Hf| > M_\Phi w) &\geq \sum_{k=0}^{L} w(B_k^o) \\
&= \frac{1}{2}\sum_{k=0}^{L} w(B_k) \\
&\geq \frac{1}{2}\sum_{k=0}^{L}(1-\varepsilon)(1+e^{-\pi K})^{-k/2}\left(1-(1+e^{-\pi K})^{-1/2}\right) \\
&= \frac{1}{2}(1-\varepsilon)\left(1-(1+e^{-\pi K})^{-(L+1)/2}\right) \\
&= \frac{1}{2}(1-\varepsilon)\left(1-(1+L^{-1})^{-(L+1)/2}\right).
\end{aligned}$$

The latter expression is bounded away from zero. On the other hand, by (6.5), the L^1-norm of f can be arbitrarily small, by the proper choice of α. This proves that the estimate (6.2) cannot hold with any finite constant.

Acknowledgments

The author would like to thank an anonymous Referee for the careful reading of the first version of the paper and several helpful suggestions.

References

[1] BUCKLEY, S. M.: Estimates for operator norms on weighted spaces and reverse Jensen inequalities. *Trans. Amer. Math. Soc.* **340** (1993), no. 1, 253–272.

[2] BURKHOLDER, D. L.: Boundary value problems and sharp inequalities for martingale transforms. *Ann. Probab.* **12** (1984), no. 3, 647–702.

[3] CALDARELLI, M. AND RIVERA-RÍOS, I. P.: A sparse approach to mixed weak type inequalities. *Math. Z.*, to appear.

[4] CALDARELLI, M., LERNER, A. K. AND OMBROSI, S.: On a counterexample related to weighted weak type estimates for singular integrals. *Proc. Amer. Math. Soc.* **145** (2017), no. 7, 3005–3012.

[5] CHANILLO, S. AND WHEEDEN, R. L.: Some weighted norm inequalities for the area integral. *Indiana Univ. Math. J.* **36** (1987), no. 2, 277–294.

[6] CRUZ-URIBE, D. AND PÉREZ, C.: Two-weight extrapolation via the maximal operator. *J. Funct. Anal.* **174** (2000), no. 1, 1–17.

[7] FEFFERMAN, C. AND STEIN, E. M.: Some maximal inequalities. *Amer. J. Math.* **93** (1971), 107–115.

[8] LACEY, M.: An elementary proof of the A_2 bound. *Israel J. Math.* **217** (2017), no. 1, 181–195.

[9] LERNER, A. K.: On an estimate of Calderón-Zygmund operators by dyadic positive operators. *J. Anal. Math.* **121** (2013), 141–161.

[10] LERNER, A. K. AND NAZAROV, F.: Intuitive dyadic calculus: the basics. *Expo. Math.* **37** (2019), no. 3, 225–265.

[11] LERNER, A. K., NAZAROV, F. AND OMBROSI, S.: On the sharp upper bound related to the weak Muckenhoupt-Wheeden conjecture. *Anal. and PDE*, to appear.

[12] LERNER, A. K., OMBROSI, S. AND PÉREZ, C.: Sharp A_1 bounds for Calderón-Zygmund operators and the relationship with a problem of Muckenhoupt and Wheeden. *Int. Math. Res. Not. IMRN* (2008), no. 6, Art. ID rnm161, 11 pp.

[13] LERNER, A. K., OMBROSI, S. AND PÉREZ, C.: Weak type estimates for singular integrals related to a dual problem of Muckenhoupt-Wheeden. *J. Fourier Anal. Appl.* **15** (2009), no. 3, 394–403.

[14] LERNER, A. K., OMBROSI, S. AND PÉREZ, C.: A_1 bounds for Calderón-Zygmund operators related to a problem of Muckenhoupt and Wheeden. *Math. Res. Lett.* **16** (2009), no. 1, 149–156.

[15] NAZAROV, F., REZNIKOV, A., VASYUNIN, V. AND VOLBERG, A.: A Bellman function counterexample to the A_1 conjecture: the blow-up of weak norm estimates of weighted singular operators. available at arXiv:1506.04710v1.

[16] NAZAROV, F., REZNIKOV, A., VASYUNIN, V. AND VOLBERG, A.: On weak weighted estimates of the martingale transform and a dyadic shift. *Anal. PDE* **11**, (2018), no. 8, 2089–2109.

[17] NAZAROV, F. AND TREIL S. R.: The hunt for a Bellman function: applications to estimates for singular integral operators and to other classical problems of harmonic analysis. *St. Petersburg Math. J.* **8** (1997), no. 5, 721–824.

[18] NAZAROV, F., TREIL S. R. AND VOLBERG, A.: The Bellman functions and two-weight inequalities for Haar multipliers. *J. Amer. Math. Soc.* **12** (1999), no. 4, 909–928.

[19] Osękowski, A.: Sharp martingale and semimartingale inequalities. *Monografie Matematyczne.* **72**, Birkhäuser Basel, 2012.

[20] Osękowski, A.: Survey article: Bellman function method and sharp inequalities for martingales. *Rocky Mountain J. Math.* **43** (2013), no. 6, 1759–1823.

[21] Osękowski, A.: A weighted weak-type bound for Haar multipliers. *Royal Soc. Edinburgh Proc A (Mathematics)* **148A** (2018), no. 3, 643–658.

[22] Pérez, C.: Weighted norm inequalities for singular integral operators. *J. Lond. Math. Soc.* **49** (1994), no. 2, 296–308.

[23] Reguera, M. C.: On Muckenhoupt-Wheeden conjecture. *Adv. Math.* **227** (2011), no. 4, 1436–1450.

[24] Reguera, M. C. and Thiele, C.: The Hilbert transform does not map $L^1(Mw)$ to $L^{1,\infty}(w)$. *Math. Res. Lett.* **19** (2012), no. 1, 1–7.

Received ??

Faculty of Mathematics, Informatics and Mechanics
University of Warsaw
Banacha 2, 02-097 Warsaw
Poland

The research was supported by National Science Centre Poland grant DEC-2014/14/E/ST1/00532.